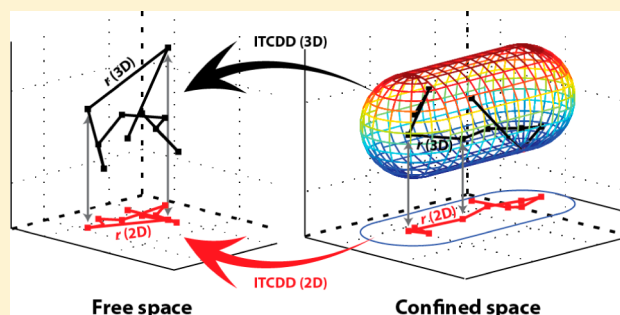Article

Quantifying Multistate Cytoplasmic Molecular Diffusion in Bacterial Cells via Inverse Transform of Confined Displacement Distribution

Tai-Yen Chen,[†] Won Jung,[†] Ace George Santiago, Feng Yang, Łukasz Krzemiński, and Peng Chen*

Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States

Ⓢ Supporting Information

ABSTRACT: Single-molecule tracking (SMT) of fluorescently tagged cytoplasmic proteins can provide valuable information on the underlying biological processes in living cells via subsequent analysis of the displacement distributions; however, the confinement effect originated from the small size of a bacterial cell skews the protein's displacement distribution and complicates the quantification of the intrinsic diffusive behaviors. Using the inverse transformation method, we convert the skewed displacement distribution (for both 2D and 3D imaging conditions) back to that in free space for systems containing one or multiple (non)interconverting Brownian diffusion states, from which we can reliably extract the number of diffusion states as well as their intrinsic diffusion coefficients and respective fractional populations. We further demonstrate a successful application to experimental SMT data of a transcription factor in living *E. coli* cells. This work allows a direct quantitative connection between cytoplasmic SMT data with diffusion theory for analyzing molecular diffusive behavior in live bacteria.

1. INTRODUCTION

Diffusive behaviors of membrane and cytoplasmic molecules in cells carry valuable information on the underlying biological processes, such as membrane protein oligomerization,[1] protein−membrane interactions,[2] protein−DNA interactions,[3] DNA repair,[4] cytokinesis,[5] and chromosome diffusion.[6] Because these processes fulfill many cellular functions, quantifying the diffusive behaviors of these molecules is important for understanding the underlying mechanisms.

A number of techniques have been developed to study the diffusive behaviors of membrane and cytoplasmic molecules. Fluorescence recovery after photobleaching (FRAP),[7] fluorescence correlation spectroscopy (FCS),[8] and single-molecule tracking (SMT)[9] are the three most common fluorescence-based methods.[10] Both FRAP and FCS probe molecular diffusive behaviors within a small volume defined by the laser focus; however, the slow time resolution and potential DNA damage caused by photobleaching in FRAP,[11] the susceptibility to optical aberrations in FCS,[12] and the diffraction-limited spatial resolution constrain the application of FRAP and FCS to molecular diffusions in live cells. On the other hand, recent technological advances in camera, fluorescent protein (FP) reporters, and super-resolution imaging algorithm[13] made it possible to track individual molecules with high spatial (few nanometers) and temporal (microseconds) resolution[14] in live cells.[15] Imaging one molecule at a time typically is through imaging a fluorescent tag, which is often a regular or photoconvertible FP. Even though the photobleaching of the fluorescent tag limits the observation time, recent studies have shown that SMT is particularly powerful in dissecting the

mechanisms of biophysical processes.[16,17] Using probes such as quantum dots or plasmonic nanoparticles can further extend SMT trajectories in time.[18]

Through real-time SMT, one directly obtains the diffusive behavior of each fluorescently labeled protein molecule in the cell reflected by its location versus time trajectory. Quantitative methods to analyze the SMT trajectories include mean-squared displacement (MSD), hidden Markov modeling (HMM),[19−22] and probability distribution function (PDF) or cumulative distribution function (CDF) of displacement length analyses. MSD analysis, the most popular method, reliably determines the diffusion coefficient for molecules moving in free space with a single diffusion state.[23] For molecules having transient diffusive behaviors or those containing multiple diffusion states, MSD method is less ideal due to its requirement of averaging over all displacements.[24] HMM analysis, a probabilistic maximum-likelihood algorithm, can extract the number of diffusion states and their interconversion rate constants (with certain assumptions);[21,22,25] it provides a mathematically derived routine and unbiasedly analyses SMT trajectories, but the resulting multistate diffusion model often lacks a definitive number of states.[26] The HMM analysis of SMT trajectories is further constrained by the complex computational algorithm and the difficulty in incorporating the photophysical kinetics of the fluorescent probe. Analysis of the PDF or CDF of displacement length on the basis of Brownian diffusion model

is known to be a robust way to quantify the diffusion coefficients and fractional populations of multistate systems, as demonstrated both in vitro and in vivo,[3−5,27−29] even though it requires more control experiments and elaborate analysis based on a defined kinetic model to extract the minimal number of diffusion states and their interconversion rate constants.

One factor that significantly affects the PDF or CDF analysis of cytoplasmic diffusion displacement is the confinement by the cell volume, especially for bacterial cells, which are less than a few microns in size. This confinement distorts and compresses the displacement length distribution, especially for molecules with large diffusion coefficients. SMT trajectories obtained from cells with different geometries can give significantly biased displacement length distributions, even though the underlying diffusion coefficient is the same. As a result, fitting the distribution of displacement length with PDF or CDF derived from the Brownian diffusion model (or any other model) only reports apparent diffusion coefficients, which are typically smaller than the intrinsic diffusion coefficients.

For membrane protein diffusion, it is a two dimension (2D) diffusion on a surface curved in three dimension (3D) space, and it does not actually have boundary confinement, as the cell membrane is a continuous boundary-less surface; however, SMT trajectories are generally obtained in 2D, where only the x, y movements in the imaging plane are tracked, thus projecting the boundary-less movements of membrane protein diffusion into a 2D diffusion confined by the cell boundary. This confinement effect from 2D projection of membrane diffusion distorts and compresses the displacement length distribution as well. To address this projection-induced confinement effect, Peterman and coworkers introduced the inverse projection of displacement distribution (IPODD) method[30] in analyzing simulated one-state membrane diffusion in bacterial cells (e.g., E. coli). In short, they first created a projected displacement distribution (PDD) matrix for a given cell geometry by projecting the simulated membrane displacement vectors onto the 2D imaging plane. For *each* displacement length that could occur anywhere on the membrane surface, they determined the resulting distribution of displacement length after projection. The PDD matrix thus quantifies the relationship from the displacement distribution before projection to that after projection. Using inverse transformation, they could then convert the 2D-projected displacement length distribution (which is often the one determined experimentally) into a most probable displacement length distribution on the cell membrane, which is readily analyzed to give the intrinsic diffusion coefficient.

Here we report an extension of the inverse transformation method for membrane diffusion to analyze cytoplasmic molecular diffusions. Using simulated diffusion trajectories in free and confined spaces, we demonstrate this inverse transformation method in analyzing 1-state cytoplasmic Brownian diffusions in both 2D and 3D and with varying diffusion coefficients and cell geometries. We further extend this method to multistate cytoplasmic diffusions, containing noninterconverting or interconverting states, to effectively extract the minimal number of diffusion states as well as their respective diffusion coefficients and fractional populations. Finally, we demonstrate a successful application to experimental SMT data of a transcription factor in living E. coli cells, which shows interconverting multistate diffusive behaviors.

2. METHODS

2.1. Simulations of Single-Molecule Diffusion Trajectories.
On the basis of the Brownian diffusion model, we used home-written Matlab codes to simulate 3D single-molecule diffusion trajectories that contained one, three noninterconverting, or three interconverting diffusion states in both free space and confined space. Each simulation condition contained at least 100 000 diffusion trajectories to ensure statistically saturated data for analysis. The 2D diffusion trajectories were generated from the 3D ones by discarding the z-component.

Diffusion Trajectories in Free Space. The 3D diffusion trajectories in free space containing one diffusion state were simulated via the following steps. First, we randomly sampled the initial position (x, y, z) in free space, where the values of x, y, and z are each from a randomly generated number. Second, with the input diffusion coefficient D we generated the distribution of displacement vector ($\vec{r}_i$, where $i = x$, y, or z) following Brownian diffusion in free space as described by eq 1, where $n = 1$, for each of the three dimensions (i.e., x, y, and z) and using a time resolution $t = 4$ or 60 ms. Third, we randomly chose a $\vec{r}_i$ from the distribution of the displacement vector, together with the initial position, to calculate the subsequent position, which also served as the new initial position for the next simulation step. The procedure was then repeated until the length of the final moving trajectory contained 10 positions for analysis. Trajectories for three noninterconverting states were generated as that in single diffusion state case but with D of 0.036, 0.7, and 11 $\mu m^2\ s^{-1}$, separately.

The 3D diffusion trajectories that contained three interconverting diffusion states were simulated with three input diffusion coefficients D_i ($i = 1$, 2, or 3) and their associated interconversion rate constants (e.g., rate constant γ_{ij} for interconversion from state i to j; $i \neq j$ and $i, j = 1$, 2, or 3). A sequence of residence time on the diffusion state i was built, where each residence time t_i sampled the residence time distribution $\exp(-\sum_j \gamma_{ij} t)$, where Σ_j was a sum of all competing processes leaving from state i to state j ($j \neq i$), each with a rate constant γ_{ij}. The transition from state i to a particular state j followed the relative probability ($\gamma_{ij}/\sum_j \gamma_{ij}$). The residence time sequence was terminated by t_{bl}, which equaled the sum of all residence times in the sequence, and t_{bl} samples the distribution $\exp(-k_{bl}(T_{int}/T_{tl})t)$, which was limited by the photobleaching and photoblinking of the fluorescence tag, where k_{bl} is the tag's intrinsic photobleaching and photoblinking rate constant. T_{int} and T_{tl} are the laser exposure time and stroboscopic imaging lapse time, respectively. During each state, the generation of displacements was the same as described in the one diffusion state case. Here we first generate the primary diffusion trajectories with T_{int} and T_{tl} of 4 ms. For trajectories with longer T_{tl}, the primary diffusion trajectories were resimulated and sampled at every lapse time T_{tl} to give the eventual simulated diffusion trajectory, which is analyzed.

Diffusion Trajectories in Confined Space. To mimic the 3D SMT data in a bacterium cell, we first modeled the 3D cell geometry as a cylinder capped by two hemispheres for simplicity with cell length and width adapted from our experimental results. The 3D diffusion trajectories in a confined space (i.e., the cell volume) were generated by similar procedures as described in free space but with random selection of initial positions inside the cell volume and the implementation of confinement effect with the boundary reflection from the cell surface. Boundary reflection was
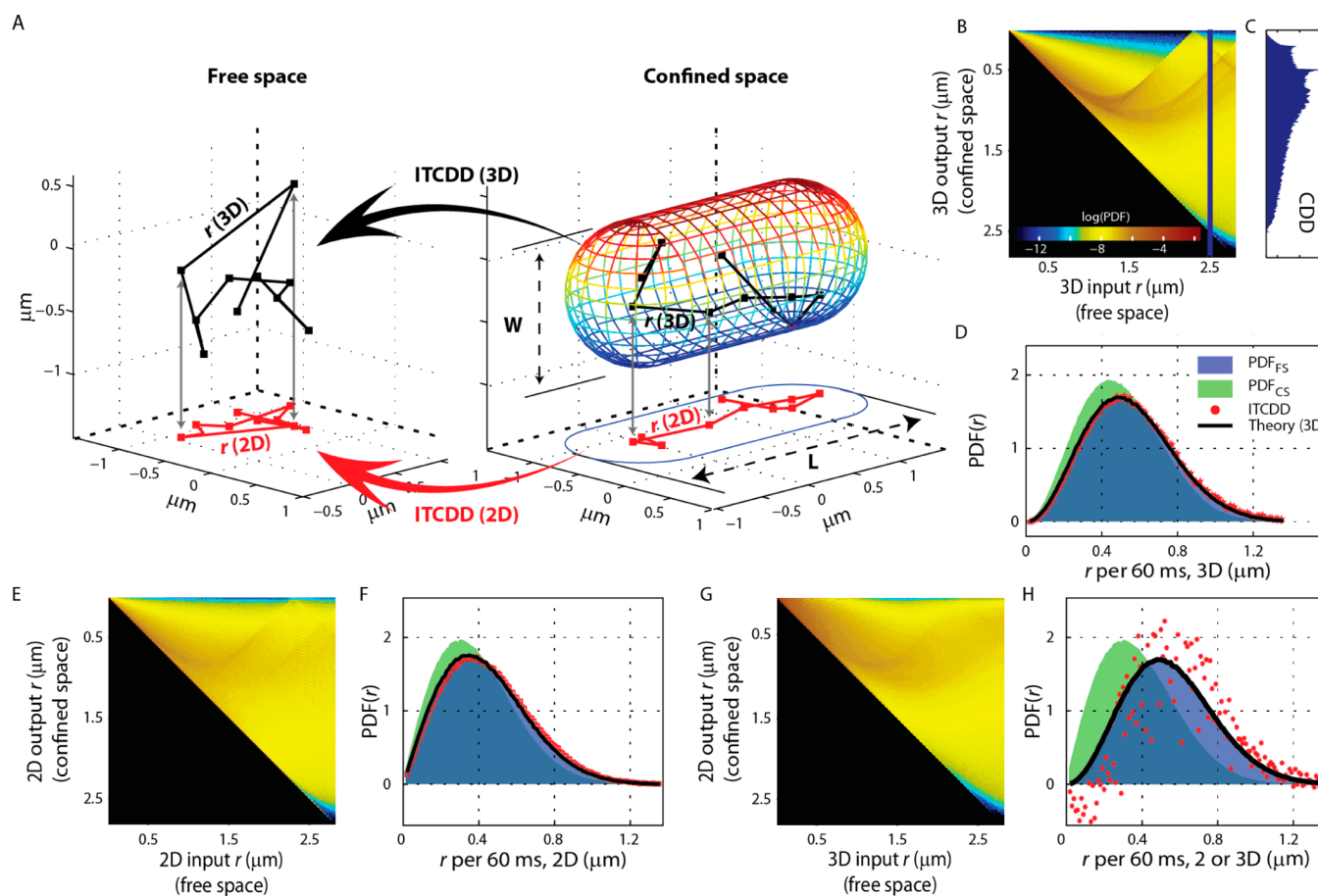
Figure 1. Illustration of inverse transform of confined displacement distribution (ITCDD) using simulated Brownian diffusions. (A) Schematic overview of ITCDD. Single-molecule diffusion trajectories are first generated in 3D in free or confined space (black trajectories) with $t = 60$ ms. Removing the z component from 3D trajectories results in the corresponding projected 2D trajectories (red trajectories). Converting the displacement length distribution in confined space to that in free space is achieved via inverse transformation of confined displacement distribution using the confinement transformation matrix ($[CTM]$). Here all confined diffusion simulations were performed within a cell having width (W) and length (L) of 1.15 and 2.82 μm, respectively. (B) $[CTM]$ for the 3D output displacements in confined space given 3D input displacements in free space. The input r is from 10 to 2820 nm with 10 nm increment. (C) CDD from B at 3D input displacement with length of 2.5 μm. (D) Overlay of simulated displacement length distributions in 3D free space (PDF_{FS}, blue shade) and in confined space (PDF_{CS}, green shade). Apply ITCCD on PDF_{CS} recovers the displacement length distribution (red symbols) that agrees well with that in free space. Both the simulated PDF_{FS} and ITCDD match the theoretical displacement length distribution (black line) of the Brownian diffusion model. All distributions are normalized with the integrated area being one. (E,F), same as panels B and D but for 2D case. (G, H) Same as panels B and D but the $[CTM]$ is from 3D input displacement to 2D output displacement. ITCDD (red dots) clearly deviates from the theoretical displacement length distribution from the Brownian diffusion model (black line).

performed when the end point of displacement vector is outside the cell volume. The intercept of cell boundary and displacement vector, together with the normal plane, was calculated for subsequent evaluation of the reflected position. The corresponding 2D simulated data were then generated from 3D ones by discarding the diffusion information in the z direction. The 3D diffusion trajectories for systems with three interconverting diffusion states in confined space were simulated in the same way as in free space but with applied boundary reflection in the displacement generation step.

2.2. Generation of Confinement Transformation Matrix. Generation of the confinement transformation matrix ($[CTM]$) was inspired by Peterman's work on inverse projection of displacement distributions (IPODD) for analyzing membrane proteins diffusing on the curved surface. In short, >100 000 displacement vectors ($\vec{r}$) of a given distance length r were randomly positioned in the cell. If the end point of displacement vector was outside the cell volume, the

boundary reflection was performed, generating final positions. We then calculated the output r from the final positions and created the confined displacement distribution (CDD), which served as a single column data for the $[CTM]$. The length of displacement vector varied from 10 nm to 2.82 μm (i.e., up to the cell length) with 10 and 30 nm increments for transforming simulated and experimental data, respectively. Finally, CDDs for all input displacement vectors were combined to form the confinement transformation matrix.

2.3. Generation of Probability Density Function of Displacement Length for Systems with Multi Diffusion States. All probability density functions (PDFs) of displacement length ($PDF(r)$) in this study were generated from the distribution of displacement length of moving trajectories normalized by the area of distribution. For example, for systems with a single diffusion state, displacements were calculated from the moving trajectory and used to generate the histogram of displacement length for a given bin size (i.e., 10 and 30 nm for

simulated and experimental data, respectively). The displacement histogram was then divided by its area to create the PDF of displacement length, PDF(r).

PDF(r) for systems with static three diffusion states was obtained as follows. We combined displacements from the respective diffusion states with given weighting coefficients to generate the displacement length histogram, which was then normalized by its area to create the PDF(r) for analysis. For example, after simulating 100 000 trajectories (with trajectory length of 10 positions) in a given cell geometry for each of three different D_{input}, we randomly chose trajectories from each diffusion state and combined them with chosen fractional populations for subsequent analyses.

Finally, for system with three interconverting diffusion states, the PDF(r) was generated from moving trajectories based on procedures as described in the Methods section. Because the moving trajectories were simulated with three interconverting diffusion states built-in, the PDF(r) was simply the resulting displacement histogram normalized by the histogram area.

2.4. Transformation of Distribution of Displacement Length between Free and Confined Spaces. Transformation of distribution of displacement length between free and confined spaces was achieved via the confinement transformation matrix ([CTM]). Forward converting the 2D or 3D distribution of displacement length in free space to that in confined space was via direct multiplication of the 2D or 3D distribution in free space with [CTM]. As for the inverse transformation (i.e., distribution in confined space to that in free space) process, the inverted [CTM] (i.e., [CTM]$^{-1}$) was first obtained using Gaussian elimination; multiplication of the distribution of displacement length in confined space with the [CTM]$^{-1}$ (i.e., eq 4) then resulted in the corresponding distribution in free space. Note that in the [CTM]$^{-1}$ obtaining step we first diagonalized the [CTM] and back-substituted the known variables to solve for [CTM]$^{-1}$ rather than simply transpose the [CTM].

3. RESULTS AND DISCUSSIONS

3.1. Inverse Transform of Confined Displacement Distribution for Cytoplasmic Molecules. The diffusive motions of cytoplasmic molecules in a bacterial cell are significantly confined by the small cell size (Figure 1A, right) (a typical *E. coli* cell is about $0.5 \times 0.5 \times 2$ μm^3 in size (e.g., ~ 1.5 fL)), and for a small protein with a diffusion coefficient of 10 μm^2 s^{-1}, its diffusion can traverse the cell length in $\sim$100 ms. This confinement effect distorts the molecule's displacement distribution, hindering the quantification of its diffusion coefficient. For heterogeneous diffusion where multiple diffusion states are present, this confinement effect also hinders the determination of the (minimal) number of diffusion states. Here we present an inverse transform method to analyze displacement distributions of confined diffusions to obtain displacement distributions that are well-described by Brownian diffusion in free space. The feasibility of the method is examined by diffusion simulations in free and confined spaces.

For 3D Brownian diffusion in free space, the probability density distribution of displacement vector $\vec{r}$ within time t follows a Gaussian function

$$P(\vec{r}, t) = \left(\frac{1}{\sqrt{4\pi Dt}} \right)^n \exp\left(-\frac{\vec{r}^2}{4Dt} \right) \quad (1)$$

where D is the diffusion coefficient and n = 1, 2, or 3 for 1D, 2D, or 3D diffusion, respectively. The second moment of $\vec{r}$ follows the well-known relationship $\langle \vec{r}^2 \rangle = 2nDt$. The PDF of the 3D and 2D displacement length, r, which is the scalar component of $\vec{r}$, can be obtained by integrating $P(\vec{r}, t)$ over all angular spaces

$$PDF(r, t) = \frac{r^2}{\sqrt{4\pi} (Dt)^{3/2}} \exp\left(-\frac{r^2}{4Dt} \right), \quad (3D) \quad (2)$$

$$PDF(r, t) = \frac{r}{2Dt} \exp\left(-\frac{r^2}{4Dt} \right), \quad (2D) \quad (3)$$

The blue shade in Figure 1D shows the distribution of displacement length r from a simulation of 3D Brownian diffusion in free space with D = 1 μm^2 s^{-1} and t = 60 ms (simulation details in the Methods section), which is well-described by eq 2 (Supplementary Figure S1). As most of the SMT experiments are done in 2D imaging mode, the blue shade in Figure 1F presents the distribution of the corresponding displacement length r in 2D, which is again well-described by eq 3 (Supplementary Figure S1). When the same Brownian diffusion is simulated inside a confined space (e.g., inside a bacterial cell, Figure 1A, right), the distributions of displacement length r in both 3D and 2D are significantly distorted due to reflections by cell boundaries (Figure 1D,F), as expected. These confined displacement length distributions do not follow eqs 2 and 3, and attempted fitting gives the diffusion constant of 0.76 ± 0.32 μm^2 s^{-1}, underestimated from the expected diffusion coefficient of 1 μm^2 s^{-1}.

To numerically mimic the confinement effect on the displacement length distribution, we followed Peterman et al.[30] to generate a confinement transformation matrix ([CTM]; e.g., Figure 1B) for a given cell geometry, which is readily measured for bacterial cells. For each column of this matrix, a 3D displacement vector in free space of a given length is randomly sampled within the cell volume and applied boundary reflections when the vector impinges on the cell boundary. In this way, it generates a distribution of corresponding 3D displacement in the confined space. Normalizing this distribution gives the CDD, which represents the probability distribution for finding a 3D-confined displacement length given a 3D displacement of a particular length in free space (Figure 1C). Varying the length of the 3D displacement vector in free space and repeating the random sampling process generates the data for all other columns in [CTM] (Figure 1B). The utility of this confinement transformation matrix can be seen by applying it to the distribution of displacement length from the simulated 3D Brownian diffusion in free space as [CTM]·PDF$_{FS}$ = PDF$_{CS}$, where PDF$_{FS}$ and PDF$_{CS}$ are PDFs of displacement length in free and confined spaces, respectively. The resulting distribution from this forward transformation reproduces that from the simulations in the confined space (Supplementary Figure S2A).

More useful is the inverse transformation of the confined displacement distribution (ITCDD), as the CDD is what is directly measured in experiments

$$PDF_{FS} = [CTM]^{-1} \cdot PDF_{CS} \quad (4)$$

where [CTM]$^{-1}$ can be obtained by Gaussian elimination (Section 2.4). Applying ITCDD on the simulated results in the confined space deconvolutes the confinement effect and effectively reproduces the theoretical distribution of displace-

ment length r in free space (Figure 1D). Fitting the inverse transformed distribution gives $D = 1.04 \pm 0.01$ μm^2 s^{-1}, reliably recovering the expected diffusion coefficient ($D = 1$ μm^2 s^{-1}). All fittings of the ITCDD were done via least-squares fitting in MATLAB program using PDFs with single or three diffusion states (e.g., eqs 3 and 5, respectively, for the 2D diffusion cases).

The forward transformation, and more importantly, the inverse transformation using the confinement transformation matrix are equally applicable between the 2D displacement length distribution in free space and that in confined space (Figure 1E,F and Supplementary Figure S2B).

It is important to point out that these forward and inverse transformations only work well when the confinement transformation matrix is generated when the input and output displacements match in dimension. Figure 1G shows the $[CTM]$ generated between 3D displacement in free space and the 2D displacement in confined space. Using this $[CTM]$ or $[CTM]^{-1}$ for forward or inverse transformation cannot reproduce the expected distributions (Figure 1H and Supplementary Figure S2C). It is worth noting that the original Peterman's work on membrane diffusion is between 2D diffusion in curved surface and its 2D projection onto a flat surface,[30] where the displacement dimensions are matching. A likely reason for the inapplicability of transformation between different dimensions is that the lower dimension displacements are missing information about the third dimension; this missed dimension cannot be created during the transformation to the higher dimension displacements.

3.2. Analysis of One Diffusion State in Cells: Variation in Diffusion Coefficient and Cell Geometry.
The diffusion coefficient (D) of cytosolic molecules in bacteria typically ranges from 10^{-2} to 10 μm^2 s^{-1}.[31] To probe whether the magnitude of the diffusion coefficient may affect the performance of the inverse transformation method, we simulated diffusion trajectories with variable D values. Because most of SMT experiments are done in 2D, we focus discussions and analyses on the 2D displacements generated from simulated diffusions that are always done in 3D; the results apply equally to the 3D displacements.

Figure 2A shows the simulated 2D PDF(r) values in free space and in a confined cell volume with the input D of 11 μm^2 s^{-1}, a typical diffusion coefficient for a fast-diffusing small protein in bacterial cytoplasm, for which the confinement effect is more significant than those with smaller diffusion coefficients. The corresponding ITCDD closely mimics that in free space (Figure 2A); fitting it with eq 3 gives D_{fit} of 12.5 ± 0.2 μm^2 s^{-1}, within 14% of the input D. With the input D varying from 0.01 to 11 μm^2 s^{-1}, the fitted D from ITCDD is always within $0.1-14\%$ of the input D (Figure 2B), smaller than or comparable to typical experimental uncertainties ($8-25\%$).[3,27] Therefore, the inverse transformation method allows for direct and reliable extraction of the intrinsic diffusion coefficients of Brownian diffusions in confined space.

Using a fixed input D (e.g., 11 μm^2 s^{-1}), we further evaluated how the cell geometry, which the $[CTM]$ is dependent on, might affect the performance of the inverse transformation method. We examined Brownian diffusions in cells with width of 1.15 μm and lengths of 2.32, 2.82, and 3.32 μm, corresponding to aspect ratios of 2.0, 2.5, and 2.9, respectively. These geometries cover the range of *E. coli* cell shapes typically observed in minimum growth medium.[3] Regardless of the cell geometry, the fitted D from ITCDD stays at 12.4 ± 0.5 μm^2 s^{-1}, within 13% of the input D (Figure 2C). To further test the
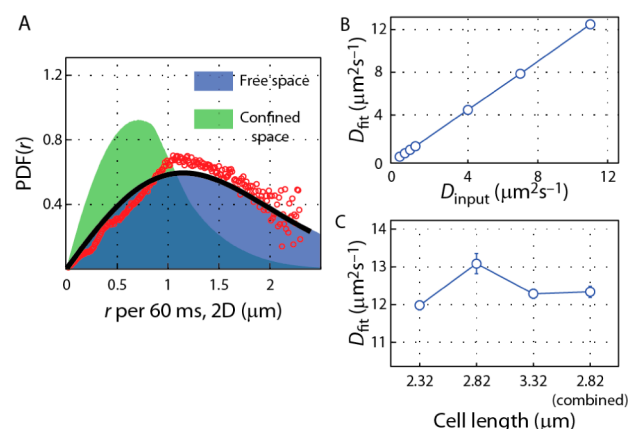

Figure 2. Analysis of one diffusion state inside a cell with ITCDD. (A) Simulated PDF(r) data of diffusion trajectories in free (blue shade) and confined (green shade) spaces with D_{input} of 11 μm^2 s^{-1} and $t = 60$ ms. Multiplication of PDF(r) of diffusion trajectories in confined space with $[CTM]^{-1}$ resulted in the corresponding ITCDD (red circles), which reproduces the theoretical distribution (black curve) based on the Brownian model. (B) Fitted diffusion coefficients of ITCDD at various input diffusion coefficients. (C) Fitted diffusion coefficients of ITCDD (D_{input} of 11 μm^2 s^{-1}) with cell lengths varying from 2.32 to 2.82 μm. The fit result of the combined ITCCD with the average cell length of 2.82 μm is also plotted for comparison.

insensitiveness of the inverse transformation method to cell geometry within this range, we combined the simulated distributions of displacement lengths from these three different geometries but applied merely the $[CTM]$ from the cells of length = 2.82 μm, which is about the average of the three cell lengths; fitting the ITCDD with eq 3 again gives an D_{fit} of 12.3 ± 0.2 μm^2 s^{-1}, close to the input D. Therefore, even when diffusion trajectories are collected from a population of cells that differ in geometry (within the range evaluated here), it is sufficient to use the $[CTM]$ for the average cell geometry to perform ITCDD to extract the intrinsic diffusion coefficient.

3.3. Analysis of Noninterconverting Multistate Diffusions in Cells.
Inside a cell, a protein molecule may have a few different diffusive behaviors depending on its interactions with other proteins, RNA, or DNA. We thus evaluated the inverse transformation method in analyzing diffusion trajectories that contain multiple (i.e., three) Brownian diffusion states. We first examined the case that these states do not interconvert, that is, a static mixture of diffusive behaviors. We again focus on the analysis of the 2D displacements here from 3D simulations. The diffusion coefficients (D_{input}) of diffusion states were set to 11 (D_1), 0.7 (D_2), and 0.036 (D_3) μm^2 s^{-1}, respectively, close to those we previously measured for a transcription factor in *E. coli* cells,[3] and the fractional population (A_3) of the D_3 state was varied from 5 to 33% while the other two fractional populations were set as $A_1 = A_2 = (1 - A_3)/2$.

Figure 3A shows the 2D PDF(r) from such a three-state simulation in a cell volume. The corresponding ITCDD can be fitted using a linear combination of PDF(r) (eq 5), each accounting for one diffusion state with its corresponding fractional population (A) as a weighting coefficient
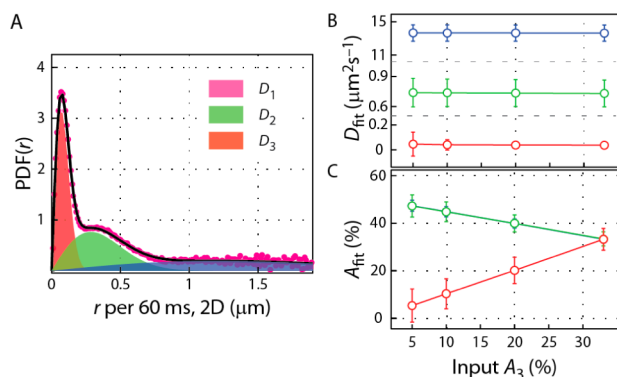
Table 1. Simulation Input Parameters of Three Interconverting Brownian Diffusion States in a Cell of 1.15 × 1.15 × 2.82 μm^3 in Size

Parameters	Simulation inputs	Kinetic scheme
D_1 ($\mu m^2 s^{-1}$)	11	
D_2 ($\mu m^2 s^{-1}$)	0.7	
D_3 ($\mu m^2 s^{-1}$)	0.036	
γ_{12} (s^{-1})	5.3	
γ_{21} (s^{-1})	2.5	
γ_{13} (s^{-1})	11.4	
γ_{31} (s^{-1})	9.6	
k_{bl} (s^{-1})	252	
Theoretic A_1 (%)	23.2	
Theoretic A_2 (%)	49.2	
Theoretic A_3 (%)	27.6	

Figure 3. Analysis of noninterconverting multistate diffusions inside a cell with ITCDD. (A) ITCDD (magenta dots) from the simulated PDF(r) of diffusion trajectories in confined space (cell width and length of 1.15 and 2.82 μm) with three noninterconverting diffusion states with D_{input} of 11 (D_1), 0.7 (D_2), and 0.036 (D_3) μm^2 s^{-1} and fractional populations of 33.3, 33.3, and 33.3%, respectively, and $t = 60$ ms. The overall fit with eq 5 (black curve) and corresponding deconvoluted three diffusion states (blue, green, and red shades for D_1, D_2, and D_3 states, respectively) were overlaid. (B) Fitted diffusion coefficients of the three diffusion states from ITCDD (blue, green, and red circles are for D_1, D_2, and D_3 states, respectively) when A_3 varies from 5 to 33%. Note that the D_{fit} of each diffusion state was plotted in different y scales for clarity. (C) Fitted fractional populations of A_1 (blue circles), A_2 (green circles), and A_3 (red circles) from ITCDD at various A_3 inputs. Note the blue and green circles are on top of each other.

$$\text{PDF}_{3state}(r, t) = \frac{A_1 r}{2D_1 t} \exp\left(-\frac{r^2}{4D_1 t}\right) + \frac{A_2 r}{2D_2 t} \exp\left(-\frac{r^2}{4D_2 t}\right)$$
$$+ \frac{(1 - A_1 - A_2)r}{2D_3 t} \exp\left(-\frac{r^2}{4D_3 t}\right) \quad (5)$$

The fitted D values of all three states are all within 10% error of D_{input} (Figure 3B). The fitted fractional populations are also in agreement with input ones (within 7% error), showing a clear trend of increasing A_3 and decreasing A_1 and A_2, as expected. Therefore, the inverse transformation method can effectively extract the intrinsic diffusion coefficients and their fractional populations of noninterconverting multistate diffusions.

3.4. Analysis of Interconverting Multistate Diffusions in Cells. Following the above section, we further evaluated the inverse transformation method in analyzing diffusion trajectories that contain three interconverting Brownian diffusion states. We simulated the 3D diffusion trajectories with 4 ms time resolution in a cell volume using a set of interconversion kinetic rate constants (Methods section) from our previous SMT study of the transcription factor CueR, which was tagged by a photoconvertible FP mEos3.2.[32] Table 1 gives input parameters of this simulation, including diffusion coefficients (D_i, $i = 1-3$) and interconverting rate constants (γ_{ij}; $i, j = 1, 2$, or 3); the interconversion rate constants also determine the fractional populations of the respective states. We further included a photobleaching rate constant (k_{bl}) to account for the fact that FP's photobleaching limits the length of tracking trajectories. Note that no interconversion was allowed between the D_2 and D_3 states because it was kinetically negligible for CueR.[3]

Focusing again on the analysis of 2D displacements, we first tested the inverse transformation method on the simulated trajectories at 60 ms time resolution (i.e., sample the

displacement from the simulated diffusion trajectories with $t = 60$ ms). Figure 4A shows the ITCDD of the 2D PDF(r) in
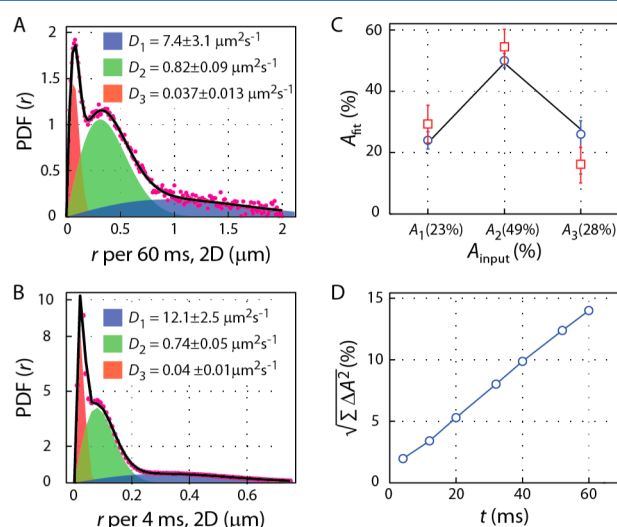

Figure 4. Analysis of interconverting multistate diffusions inside a cell with ITCDD. (A) ITCDD (magenta dots) of the simulated PDF(r) of diffusion trajectories with 60 ms time resolution in confined space with three interconverting diffusion states with D_{input} in Table 1. The overall fit result (black curve) with eq 5 and corresponding three diffusion states (blue, green, and red shades for D_1, D_2, and D_3 states, respectively) were overlaid. (B) Same as panel A but with 4 ms time resolution. (C) Fitted fractional populations (i.e., A_1, A_2, and A_3) from ITCCD with 4 (blue circles) and 60 (red squares) ms time resolution, along with the input A_1, A_2, and A_3. (D) (SSE)$^{1/2}$ of fitted fractional populations from ITCCD at various time resolutions.

confined space from the simulation. Fitting it with eq 5 gives D_1, D_2, and D_3 of 7.4 ± 3.1, 0.82 ± 0.09, and 0.037 ± 0.013 $\mu m^2 s^{-1}$ and A_1, A_2, and A_3 of 29 ± 6, 55 ± 6, and 16 ± 8.5%, respectively. Compared with the simulation inputs, the ~10% error in the fitted diffusion coefficients and fractional populations again support that the inverse transformation method can effectively deconvolute the confinement effect.

One possible reason for the ~10% error in the fitted values of diffusion coefficient and fractional population of each state is the insufficient time resolution in sampling the simulated diffusion trajectories. t of 60 ms corresponds to a sampling rate of 16.7 s^{-1}, which is comparable to the interconversion rate constant γ_{13} between D_1 and D_3 state (Table 1). We therefore also analyzed the PDF(r) by sampling the simulated diffusion trajectories at $t = 4$ ms (Figure 4B). Fitting the ITCDD of this higher time resolution results with eq 5 gives D_1, D_2, and D_3 of

12.1 ± 2.5, 0.74 ± 0.05, and 0.04 ± 0.01 $\mu m^2 s^{-1}$ and A_1, A_2, and A_3 of 24 ± 3, 50 ± 2, and 26 ± 4% respectively, which now has ~3.6% error, significantly improved compared with those in analyzing the 60 ms time resolution results. Figure 4C compares the fractional populations from fitting ITCDD of 4 and 60 ms 2D displacement lengths with A_{inputs}. The results at 4 ms resolution can perfectly recover the correct fractional populations. Therefore, as long as the displacement is obtained at sufficient time resolution in SMT measurements, the inverse transformation of CDD is effective.

What time resolution (i.e., sampling rate) would then be sufficient? To address this, we systematically varied t from 4 to 60 ms in sampling the displacements in the simulated diffusion trajectories. Figure 4D shows the square root of the sum of square error (i.e., SSE) of fractional populations (SSE = $\sum \Delta A^2$, $\Delta A = A_{fit} - A_{input}$) as a function of sampling time t. As t gets longer, $(SSE)^{1/2}$ gets larger. Assuming $(SSE)^{1/2}$ $(= (\sum \Delta A^2)^{1/2})$ < 10% as being a good fit, t of 40 ms would be a minimum time resolution here for sampling the displacements so that the inverse transformation method would give the correct fractional populations of diffusion states. This t = 40 ms, corresponding to a rate of 25 s^{-1}, is about 2.2 times faster than the fastest interconversion rate constant γ_{13} of 11.4 s^{-1} in the simulation.

3.5. Application to Transcription Regulator Dynamics in Live *E. coli* Cells. After validating the inverse transform method using simulated diffusion trajectories, we applied this method on the SMT data of CueR (in its apo form, i.e., apo-CueR), a Cu$^+$-responsive MerR-family transcription regulator, in living *E. coli* cells to extract the diffusion coefficient and fractional population of each diffusion state. Details of obtaining the SMT data were described in our previous work.[3] In short, we tagged the nonmetallated apo-CueR with the photoconvertible FP mEos3.2, generating CueR$_{apo}^{mE}$. We then used time-lapse stroboscopic imaging to track the 2D motions of individual photoconverted CueR$_{apo}^{mE}$ in a cell at a sampling rate of every 60 ms until the mEos3.2 tag photobleached.

CueR can interact with DNA specifically at recognition sites or with DNA nonspecifically.[33] Three effective diffusion states are thus expected for CueR$_{apo}^{mE}$ in an *E. coli* cell: (1) specifically bound (SB) to chromosomal recognition sites, whose diffusion coefficient should be very small and largely reflect the chromosome conformational flexibility in the cell; (2) non-specifically bound (NB) and moving along the chromosome; and (3) freely diffusing (FD) in the cytoplasm. Figure 5A shows the ITCDD from the measured 2D PDF(r) of CueR$_{apo}^{mE}$ at a low cellular protein concentration of ~100 nM. (The cellular protein concentration was quantified for each cell in our imaging approach; details see our previous work.[3]) Minimally three diffusion states are needed to fit the ITCDD satisfactorily. The three diffusion coefficients are 14.9 ± 7.6, 0.93 ± 0.08, and 0.062 ± 0.005 $\mu m^2 s^{-1}$, assignable as CueR$_{apo}^{mE}$ being FD in the cytoplasm and NB and SB to chromosome, respectively. On the contrary, at a high cellular protein concentration of ~1375 nM, the ITCDD only requires minimally two diffusion states to be fitted satisfactorily (Figure 5B). The two diffusion coefficients of 7.0 ± 1.5 and 0.90 ± 0.08 $\mu m^2 s^{-1}$ are within error to those of the FD and NB states at the low cellular protein concentration. Therefore, at this high cellular protein concentration, the SB state is no longer resolvable; this is not surprising because the fractional population of the SB state (i.e., (number of proteins specifically bound to DNA recognition sites)/(total number of proteins)) should be increasingly smaller
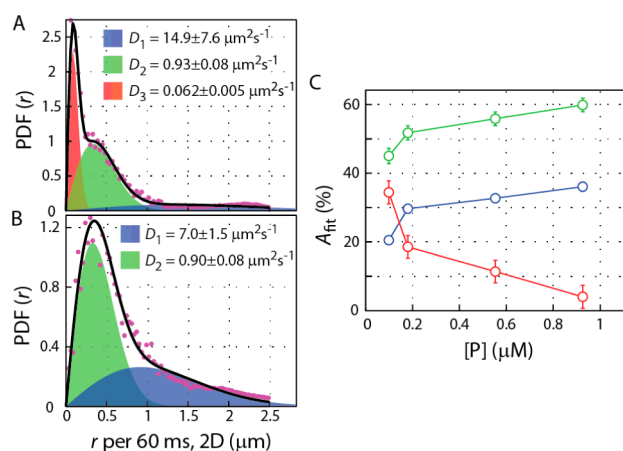

Figure 5. Analysis of diffusive behaviors of apo-CueR in *E. coli* cells with ITCDD. (A) ITCDD (magenta dots) from experimental PDF(r) data of apo-CueR with 60 ms sampling rate in cells with total cellular protein concentration ([apo-CueR]) of 99 nM. The ITCDD requires a model with three diffusion states to achieve satisfactory fitting. The overall fit result (black curve) and corresponding three resolved diffusion states (blue, green, and red shades for D_1, D_2, and D_3 states, respectively) were overlaid. (B) Same as panel A but with total [apo-CueR] = 1375 nM. Two diffusion states are sufficient to fit the ITCDD satisfactorily. (C) Fitted fractional populations of A_1, A_2, and A_3 of ITCDD for [apo-CueR] varying from 99 to 926 nM, in which a global fit across four sets of data of different cellular protein concentrations was performed. The [CTM] was based on the experimentally determined cell geometry of cell width and length of 1.15 and 2.82 μm.

at higher cellular protein concentrations. More importantly, these results demonstrate that depending on the cellular protein concentration the experimentally resolvable number of diffusion states can vary.

To more reliably determine the minimal number of diffusion states, we propose using a global fit of ITCDD across all cellular protein concentrations (i.e., four sets of data, each set coming from a sorted group of cells with a particular cellular protein concentration[3]), where the number of diffusion states and their diffusion coefficients are shared. This global fit on the results of CueR$_{apo}^{mE}$ gave three states with diffusion coefficients of 8.2 ± 0.3, 0.92 ± 0.04, and 0.051 ± 0.005 $\mu m^2 s^{-1}$, corresponding to the FD, NB, and SB states, respectively. Compared with literature values,[4,17,27,34] they are in excellent agreements with those expected for freely diffusing in the cytoplasm, nonspecifically bound to chromosome, and specifically bound to chromosome. Figure 5C summarizes the extracted fractional populations as a function of cellular protein concentration. With increasing protein concentration, the fractional population of the SB state decreases while those of NB and FD states increase, consistent with expectations and our previous study[3] and further supporting the effectiveness of globally fitting the ITCDD.

4. CONCLUSION

High spatial and temporal resolution position trajectories from SMT of fluorescently tagged cytoplasmic proteins carry valuable information about the underlying biological processes in cells, but their analysis is complicated by the confinement effect from the cell volume, especially for small bacterial cells. Here we deconvolute out the confinement effect by inverse transforming the PDF of displacement length (PDF(r)) using

the confinement transformation matrix ([CTM]), building on the previous work of Peterman[30] on simulated single-state membrane diffusions. Besides treating single-state cytoplasmic diffusions, we further extended this method to analyze multistate Brownian diffusions in the cytoplasm, including both noninterconverting and interconverting three-state diffusions. We demonstrated the effectiveness of this method in determining the minimal number of diffusion states, their diffusion coefficients, and fractional populations as well as how to choose a sufficient time resolution in analyzing systems containing interconverting multistates. A successful application to experimental multistate SMT data of a transcription factor in live *E. coli* cell is also demonstrated. Together with Peterman's early work on membrane diffusion (whose extension to multistate systems can readily follow our work on cytoplasmic diffusion here), our method allows for direct connection between SMT data with diffusion theory for analyzing molecular diffusive behaviors in live bacteria.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b08654.

> Validation of Brownian diffusion simulation and forward transformation of displacement length distribution in free space with CTM. (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: pc252@cornell.edu.

Author Contributions
[†]T.-Y.C. and W.J. contributed equally.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Leake, M. C.; Greene, N. P.; Godun, R. M.; Granjon, T.; Buchanan, G.; Chen, S.; Berry, R. M.; Palmer, T.; Berks, B. C. Variable Stoichiometry of the Tata Component of the Twin-Arginine Protein Transport System Observed by in Vivo Single-Molecule Imaging. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (40), 15376−15381.

(2) Lillemeier, B. F.; Pfeiffer, J. R.; Surviladze, Z.; Wilson, B. S.; Davis, M. M. Plasma Membrane-Associated Proteins Are Clustered into Islands Attached to the Cytoskeleton. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (50), 18992−18997.

(3) Chen, T.-Y.; Santiago, A. G.; Jung, W.; Krzeminski, L.; Yang, F.; Martell, D. J.; Helmann, J. D.; Chen, P. Concentration- and Chromosome-Organization-Dependent Regulator Unbinding from DNA for Transcription Regulation in Living Cells. *Nat. Commun.* **2015**, *6*, 7445.

(4) Uphoff, S.; Reyes-Lamothe, R.; Garza de Leon, F.; Sherratt, D. J.; Kapanidis, A. N. Single-Molecule DNA Repair in Live Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (20), 8063−8068.

(5) Niu, L.; Yu, J. Investigating Intracellular Dynamics of Ftsz Cytoskeleton with Photoactivation Single-Molecule Tracking. *Biophys. J.* **2008**, *95* (4), 2009−2016.

(6) Javer, A.; Long, Z.; Nugent, E.; Grisi, M.; Siriwatwetchakul, K.; Dorfman, K. D.; Cicuta, P.; Cosentino Lagomarsino, M. Short-Time Movement of E. Coli Chromosomal Loci Depends on Coordinate and Subcellular Localization. *Nat. Commun.* **2013**, *4*, 3003.

(7) Axelrod, D.; Koppel, D. E.; Schlessinger, J.; Elson, E.; Webb, W. W. Mobility Measurement by Analysis of Fluorescence Photobleaching Recovery Kinetics. *Biophys. J.* **1976**, *16* (9), 1055−1069.

(8) Hess, S. T.; Huang, S.; Heikal, A. A.; Webb, W. W. Biological and Chemical Applications of Fluorescence Correlation Spectroscopy: A Review. *Biochemistry* **2002**, *41* (3), 697−705.

(9) Patterson, G.; Davidson, M.; Manley, S.; Lippincott-Schwartz, J. Superresolution Imaging Using Single-Molecule Localization. *Annu. Rev. Phys. Chem.* **2010**, *61* (1), 345−367.

(10) Mueller, F.; Stasevich, T. J.; Mazza, D.; McNally, J. G. Quantifying Transcription Factor Kinetics: At Work or at Play? *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48* (5), 492−514.

(11) Solarczyk, K. J.; Zarębski, M.; Dobrucki, J. W. Inducing Local DNA Damage by Visible Light to Study Chromatin Repair. *DNA Repair* **2012**, *11* (12), 996−1002.

(12) Ries, J.; Schwille, P. Fluorescence Correlation Spectroscopy. *BioEssays* **2012**, *34* (5), 361−368.

(13) Small, A.; Stahlheber, S. Fluorophore Localization Algorithms for Super-Resolution Microscopy. *Nat. Methods* **2014**, *11* (3), 267−279.

(14) Hiramoto-Yamaki, N.; Tanaka, K. A. K.; Suzuki, K. G. N.; Hirosawa, K. M.; Miyahara, M. S. H.; Kalay, Z.; Tanaka, K.; Kasai, R. S.; Kusumi, A.; Fujiwara, T. K. Ultrafast Diffusion of a Fluorescent Cholesterol Analog in Compartmentalized Plasma Membranes. *Traffic* **2014**, *15* (6), 583−612.

(15) Fernandez-Suarez, M.; Ting, A. Y. Fluorescent Probes for Super-Resolution Imaging in Living Cells. *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (12), 929−943.

(16) Yu, J.; Xiao, J.; Ren, X.; Lao, K.; Xie, X. S. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* **2006**, *311*, 1600−1603.

(17) Elf, J.; Li, G.-W.; Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **2007**, *316*, 1191−1194.

(18) Pinaud, F.; Clarke, S.; Sittner, A.; Dahan, M. Probing Cellular Events, One Quantum Dot at a Time. *Nat. Methods* **2010**, *7* (4), 275−285.

(19) Beausang, J. F.; Zurla, C.; Manzo, C.; Dunlap, D.; Finzi, L.; Nelson, P. C. DNA Looping Kinetics Analyzed Using Diffusive Hidden Markov Model. *Biophys. J.* **2007**, *92* (8), L64−L66.

(20) Chung, I.; Akita, R.; Vandlen, R.; Toomre, D.; Schlessinger, J.; Mellman, I. Spatial Control of Egf Receptor Activation by Reversible Dimerization on Living Cells. *Nature* **2010**, *464* (7289), 783−787.

(21) Das, R.; Cairo, C. W.; Coombs, D. A Hidden Markov Model for Single Particle Tracks Quantifies Dynamic Interactions between Lfa-1 and the Actin Cytoskeleton. *PLoS Comput. Biol.* **2009**, *5* (11), e1000556.

(22) Persson, F.; Linden, M.; Unoson, C.; Elf, J. Extracting Intracellular Diffusive States and Transition Rates from Single-Molecule Tracking Data. *Nat. Methods* **2013**, *10* (3), 265−269.

(23) Michalet, X. Mean Square Displacement Analysis of Single-Particle Trajectories with Localization Error: Brownian Motion in an Isotropic Medium. *Phys. Rev. E* **2010**, *82* (4), 041914.

(24) Bosch, Peter J.; Kanger, Johannes S.; Subramaniam, V. Classification of Dynamical Diffusion States in Single Molecule Tracking Microscopy. *Biophys. J.* **2014**, *107* (3), 588−598.

(25) Robson, A.; Burrage, K.; Leake, M. C. Inferring Diffusion in Single Live Cells at the Single-Molecule Level. *Philos. Trans. R. Soc., B* **2013**, *368* (1611), 20120029.

(26) Blanco, M.; Johnson-Buck, A.; Walter, N. Hidden Markov Modeling in Single-Molecule Biophysics. In *Encyclopedia of Biophysics*; Roberts, G. K., Ed.; Springer: Berlin, 2013; pp 971−975.

(27) Gebhardt, J. C. M.; Suter, D. M.; Roy, R.; Zhao, Z. W.; Chapman, A. R.; Basu, S.; Maniatis, T.; Xie, X. S. Single-Molecule

Imaging of Transcription Factor Binding to DNA in Live Mammalian Cells. *Nat. Methods* **2013**, *10* (5), 421−426.

(28) Morisaki, T.; Müller, W. G.; Golob, N.; Mazza, D.; McNally, J. G. Single-Molecule Analysis of Transcription Factor Binding at Transcription Sites in Live Cells. *Nat. Commun.* **2014**, *5*, 4456.

(29) Bakshi, S.; Choi, H.; Mondal, J.; Weisshaar, J. C. Time-Dependent Effects of Transcription- and Translation-Halting Drugs on the Spatial Distributions of the Escherichia Coli Chromosome and Ribosomes. *Mol. Microbiol.* **2014**, *94* (4), 871−887.

(30) Oswald, F.; L. M. Bank, E.; Bollen, Y. J. M.; Peterman, E. J. G. Imaging and Quantification of Trans-Membrane Protein Diffusion in Living Bacteria. *Phys. Chem. Chem. Phys.* **2014**, *16* (25), 12625−12634.

(31) Elowitz, M. B.; Surette, M. G.; Wolf, P.-E.; Stock, J. B.; Leibler, S. Protein Mobility in the Cytoplasm Ofescherichia Coli. *J. Bacteriol.* **1999**, *181* (1), 197−203.

(32) Zhang, M.; Chang, H.; Zhang, Y.; Yu, J.; Wu, L.; Ji, W.; Chen, J.; Liu, B.; Lu, J.; Liu, Y.; et al. Rational Design of True Monomeric and Bright Photoactivatable Fluorescent Proteins. *Nat. Methods* **2012**, *9* (7), 727−729.

(33) Joshi, C. P.; Panda, D.; Martell, D. J.; Andoy, N. M.; Chen, T.-Y.; Gaballa, A.; Helmann, J. D.; Chen, P. Direct Substitution and Assisted Dissociation Pathways for Turning Off Transcription by a Merr-Family Metalloregulator. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15121−15126.

(34) Mehta, P.; Jovanovic, G.; Lenn, T.; Bruckbauer, A.; Engl, C.; Ying, L.; Buck, M. Dynamics and Stoichiometry of a Regulated Enhancer-Binding Protein in Live Escherichia Coli Cells. *Nat. Commun.* **2013**, *4*, 1997.